Proceedings Paper:

Jiang, Tao, Grace, David orcid.org/0000-0003-4493-7498 and Mitchell, Paul D orcid.org/0000-0003-0714-2581 (2009) Improvement of pre-partitioning on reinforcement learning based spectrum sharing. In: IET International Communication Conference on Wireless Mobile & Computing (CCWMC), Shanghai, China, December 2009. International Communication Conference on Wireless Mobile and Computing (CCWMC 2009), 07-09 Dec 2009 IEEE , CHN , 299 -302.

Improvement of Pre-Partitioning on Reinforcement Learning Based Spectrum Sharing

Tao Jiang, David Grace, Paul D Mitchell

Communication Research Group, Department of Electronics, University of York, York, YO10 5DD, United Kingdom
Email: {tj511|dg| pdm} @ohm.york.ac.uk

Keywords: Pre-partitioning; reinforcement learning; cognitive radio; spectrum sharing.

Abstract

In this paper, we present a novel technique for reinforcement learning based spectrum sharing schemes. It shows how the potential action space of users is reduced by randomly partitioning the spectrum for different users. Cognitive agents are therefore able to finish their exploration stage faster than more basic reinforcement learning based schemes. Simulation results show how the communication system is more stable if all users reserve a subset of resources at the beginning and then learn to select the best channels in the reserved spectrum set. The system performance by using pre-partitioning is investigated and comparison with the traditional reinforcement learning based schemes is given.

1 Introduction

Cognitive radio is considered as a more flexible and intelligent solution that has the potential to solve the conflict between the inefficient usage of spectrum and the rapid growth of wireless services [1, 2]. A cognitive radio is designed to be '*a radio that is aware of and can sense its environment, learn from its environment and adjust its operation according to some objective function*' [3]. It can be seen from its definition that the ability of learning should be a crucial part of cognitive radio. How to intelligently interact with the dynamic wireless environment is drawing increasing amounts of attention [4-7].

Reinforcement learning is a learning approach which emphasizes individual learning from direct interactions with a dynamic environment [8]. This distinct feature of reinforcement learning makes it perfectly suited to a distributed cognitive radio scenario. However, our previous research on the implementation of reinforcement learning to cognitive radio spectrum sharing shows how the inherent exploration versus exploitation tradeoff of reinforcement learning has a significant influence on system performance [9, 10]. During an exploration period, the communicating system will receive more interruptions when users are exploring their available spectrum space. Therefore, a short but efficient exploration stage is desirable.

A basic two stage reinforcement learning based spectrum sharing scheme is proposed in our previous work [10]. The first stage which is called 'preplay' is a stage where cognitive radio users search for optimum resources and learn from the experience of searching. Once users have obtained a set of preferred resources, the exploration stage is then finished. Cognitive radio users will move to the 'main' stage thereafter and only use the spectrum assigned a higher usage priority.

The purpose of this paper is to introduce a technique called pre-partitioning. The users will randomly reserve a certain amount of spectrum resources before their transmission starts. In other words, the available spectrum pool is partitioned at the beginning of the communication. The available action space which a cognitive radio needs to explore is then significantly reduced, meaning that the exploration stage is shortened consequently.

The improvement obtained by applying pre-partitioning to a reinforcement learning spectrum sharing scheme is investigated in detail, and the system performance is compared with the basic reinforcement learning based scheme used in [10] which does not use pre-partitioning. As a further comparison the strategies are compared with a no learning scheme where the cognitive radio user only uses spectrum sensing, to illustrate the level of improvement that can be achieved by different techniques.

This paper is organized as follows: in section II, we introduce the concept of pre-partitioning for reinforcement learning based cognitive radio spectrum sharing. Then simulation results are discussed in section III. Finally, conclusions are drawn in section IV.

2 Pre-partitioning for reinforcement leaning spectrum sharing

2.1 Preplay

The fundamental issue of reinforcement learning is the trade-off between exploration and exploitation [8]. To obtain enough knowledge to distinguish between the excellent and poor actions, an agent needs to take the actions repeatedly. However, to discover such actions a learning agent has to try as many different actions as possible. Neither exploration nor exploitation can be performed exclusively in the learning process. This issue has been the subject of investigation for decades in other areas.

We proposed a 'pre-play' stage and a preferred resource set for cognitive radio spectrum sharing [10]. In the 'preplay' stage, a cognitive radio explores the available spectrum pool by accessing all channels with equal probability. The weights of the used resources are updated after every action. A specific weight threshold is also defined to distinguish preferred channels. After a user has obtained a full set of preferred resources, the stage of exploring will be suspended and the user will start to exploit the preferred resources. This stage is named the 'main' stage which is effectively the exploitation stage. The user will move back to preplay again if the weight of any preferred resource has decreased under the weight threshold. The phases of exploration and exploitation are then controllable and therefore the trade-off can be more carefully controlled.

2.2 Pre-partitioning

A drawback of reinforcement learning was highlighted in our previous work [10] in that the communication system will receive more interruption when the majority of users are searching resources in the exploration stage [9, 10]. The dropping probability and the blocking probability will be significantly higher if a user stays in preplay than if they operate in the main stage. This is because the users access all physical resources with equal probability in preplay, thereby increasing the possibility that a new activation will disturb other users. However, in the main stage they will only access the best available channels. Therefore, a quick and efficient exploration is desirable particularly in a reinforcement learning based spectrum sharing scenario. Our approach to solve this problem is to randomly partition the spectrum pool into different subsets. Users will randomly select a subset of resources as their reserved channels. They will then access their reserved spectrum only.
By reserving a certain amount of resources at the beginning of the communication, the requirement for exploration is significantly reduced. The stage of exploration is shortened accordingly. The simulation results that will be shown later illustrate how the system performance of reinforcement learning based scheme can be improved further by applying the pre-partitioning.

2.3 Algorithm

The spectrum sharing algorithm is illustrated in figure 1. The main difference between the proposed pre-partitioning scheme and the basic reinforcement learning based scheme [10] is the pre-partitioning part. The remainder of the algorithm remains the same. The strategy of spectrum assignment is different in the preplay and in the main stages. This part of the process is highlighted in the flowchart in order to help readers to gain a thorough understanding.
The cognitive radio user will firstly reserve a certain amount of channels and then select appropriate spectrum to communicate according to different strategies depending on which stage the user is in. Based on the level of success, the weight of the used resource is modified and then stored in the knowledge base. This information will be utilized as guidance in selection of resource for future transmission.

2.4 Reward function

In this paper, the reward function used to modify the weight according to the success level of an action is defined as follows:

$$W_t = f_1 \cdot W_{t-1} + f_2 \qquad (1)$$

Where W_t is the weight at time t, W_{t-1} is the weight of a channel at time t-1. f_1 and f_2 are the weighting factors that reflect the level of success of the previous action.

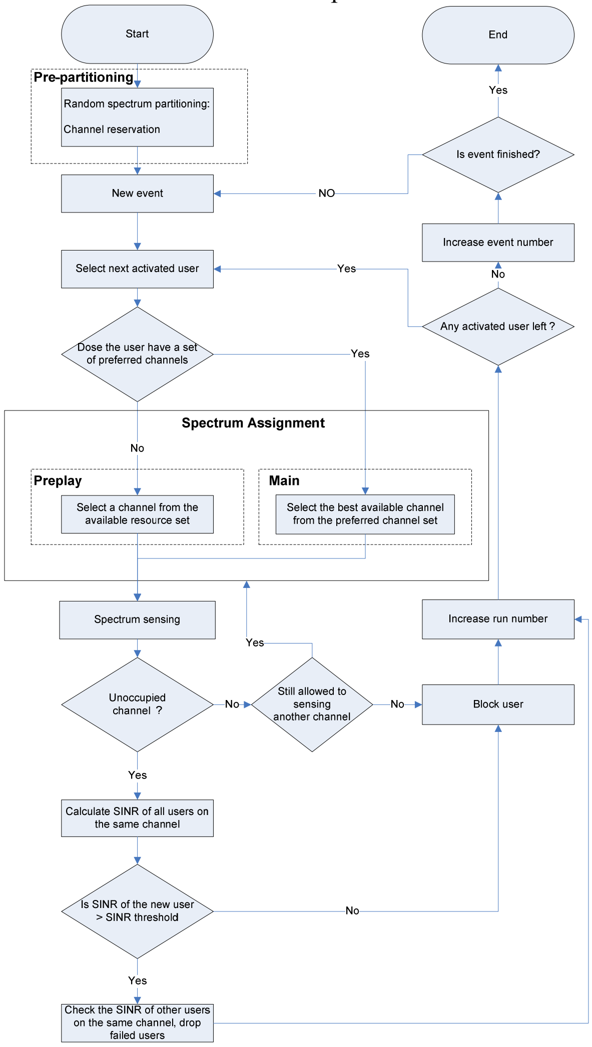

Figure 1 Flowchart of reinforcement based spectrum sensing algorithm

3 Simulation results

A simulation has been developed based on the following parameters. 1000 cognitive transmitter-receiver pairs are uniformly distributed on a square service area of $1000km^2$. The link length of the transmitter-receiver pairs is uniformly distributed between 200m-1500m. The available spectrum contains 100 channels in total. An event based scenario is applied in this simulation. A maximum number of 400 users are randomly activated at each event where the selection of the activated users follows a uniform distribution. At the beginning of the simulation, each cognitive user reserves 20 channels randomly from the spectrum pool.
The Okumura-Hata propagation model [11] is used along with log-normal shadowing with a standard deviation of 8dB. A carrier frequency of 700MHz is used and the transmitter antenna height is set to 30m. The transmitter power is fixed at 20dBm and no further power control policy is applied. The antenna gain of transmitter and receiver are fixed at 0dBi. The

interference threshold is -30dBm for spectrum sensing. The SINR threshold is set to 10dB. A noise floor of -114dBm is used, which corresponds to a noise bandwidth of 1MHz and a receiver noise temperature of 300K.

The weight is modified in exactly the same way as the restricted sensing scheme in [10], where the weight is increased by a reward of 1 if a communication request of a user is successfully accepted. Otherwise, the weight will be decreased by a punishment of 1.

A Cumulative Distribution Function (CDF) is the main mathematic tool used in this paper to analyze the simulation results. Measurements have been obtained at regular points over the service area. The CDF of the statistics of the measurements at these points is then derived. It is very important that all the parameters including the locations of users are kept exactly the same for each scheme evaluation. Thus, the different system performance is caused only by different spectrum sharing strategies.

The probability of transmission blocking and the probability of transmission dropping are the main measurements we used to evaluate system performance in this paper. The blocking probability at time t can be defined as:

$$P_B(t) = \frac{N_b(t)}{N_a(t)} \tag{2}$$

Where $P_B(t)$ is the blocking probability at time t. $N_b(t)$ is the total number of blocked activations of the system at time t and $N_a(t)$ is the total number of activations of the system at time t. similarly, the dropping probability is defined as follows:

$$P_D(t) = \frac{N_d(t)}{N_{sa}(t)} \tag{3}$$

Where $P_D(t)$ is the dropping probability at time t. $N_b(t)$ is the total number of dropped transmissions at time t and $N_{sa}(t)$ is the total number of accepted activations at time t.

The overall blocking probability of three schemes has been compared in figure 2. The red solid line represents the CDF of the blocking probability of the pre-partitioning scheme and achieves the best performance. The performance of the reinforcement learning based algorithm has been improved further by random spectrum pre-partitioning. This is because random pre-partitioning will significantly reduce the size of the available spectrum pool of each user. Therefore, the requirements for a learning part of the agent to explore the action space are reduced. In other words, the initial exploration stage of a cognitive radio user is accelerated by pre-partitioning.

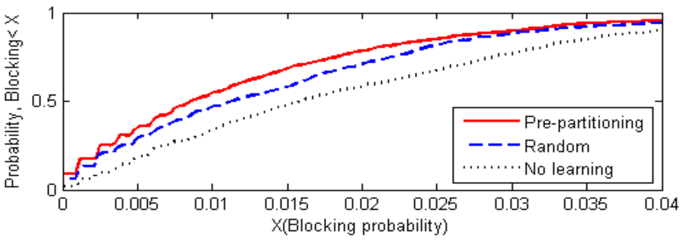

Figure 2 CDF of system blocking probability at discrete points over the service area

Figure 3 shows the CDF of dropping probability which illustrates the level of system interruption. The situation of overall dropping probability is similar to blocking probability. The scheme where users reserve a random subset of spectrum has the smallest dropping probability. The system will receive fewer interruptions if the spectrum is partitioned before the transmission starts.

The reinforcement learning spectrum sharing scheme [10] is a two stage scheme where the phases of exploration and exploitation are controllable. The preplay stage is effectively the exploration phase. Figure 4 and figure 5 illustrate the system performance in more details. The performance in different stages can be clearly seen. The system blocking probability of the pre-partitioning scheme in preplay stage is mainly higher than the basic reinforcement learning scheme. However, the red solid line outperforms the blue dash line at the lower end. It means there are about 5% users who have already reached the best performance in the communication system and will obtain an even lower blocking probability in preplay if the spectrum is pre-partitioned. These users receive less interference from the cognitive radio pairs who have the potential to interrupt their transmission, because the interfering pairs are constrained in their reserved spectrum set and no longer a source of interference. The majority of users on the contrary suffer from a higher level of blocking probability since they only have access to 20% resource of the entire spectrum pool which means they have fewer alternatives if the reserved channels are not suitable for communicating. The drawback of pre-partitioning is clear that some users may be constrained to a set of channels which have a high level of interference. Consequently these users may find it difficult to find unoccupied spectrum to use for communication.

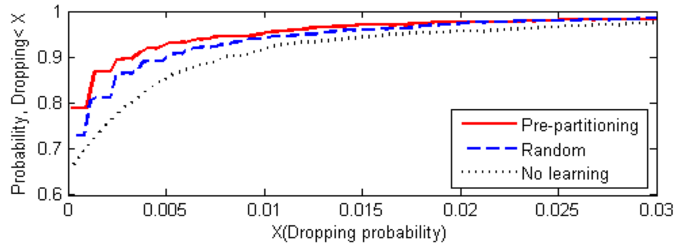

Figure 3 CDF of system dropping probability at discrete points over the service area

The pre-partitioning scheme achieves a lower blocking probability in the main stage. It can be explained by the same reason illustrated above. The system will be more stable in the main stage because the transmissions of users are restricted in the reserved channel set.

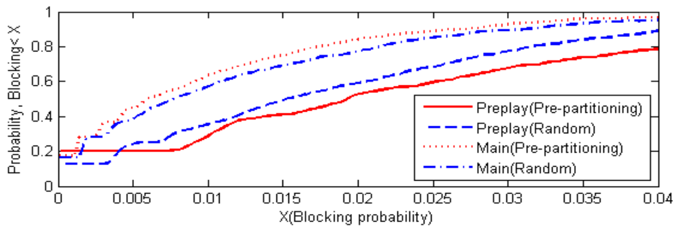

Figure 4 CDF of blocking probability in different stages at discrete points over the service area

Figure 5 shows the dropping probability of the pre-partitioning scheme and the basic reinforcement learning scheme [10] in two different stages. In the basic reinforcement learning scheme approximately 74% users are never dropped by the system at all, and this figure will increased to 83% if pre-partitioning is applied. The pre-partitioning scheme also achieves a better result in terms of dropping probability in the main stage.

It can be clearly seen from figure 2 to figure 5 that the overall performance of the pre-partitioning scheme is considerably better than the basic reinforcement learning scheme [10] even if the pre-partitioning scheme performs worse in the preplay stage. The reason can be explained by figure 6 where the details of activation probability are given. The activation probability in a certain stage can be defined as follows:

$$P_{ap} = \frac{N_{ap}}{N_{ap} + N_{am}} \tag{4}$$

$$P_{am} = \frac{N_{am}}{N_{ap} + N_{am}} \tag{5}$$

Where P_{ap} is the probability of a user to be activated in the preplay stage and P_{am} is the probability to be activated in the main stage. N_{ap} stands for the number of activations of a user in preplay through the simulation. N_{am} is the number of activations in the main stage accordingly.

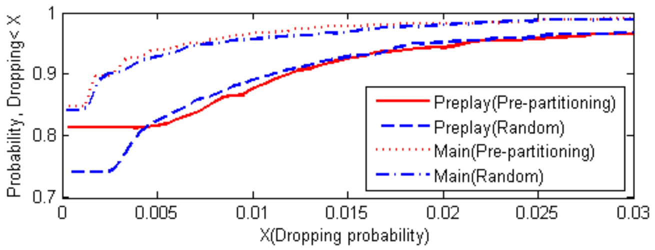

Figure 5 CDF of dropping probability in different stages at discrete points over the service area

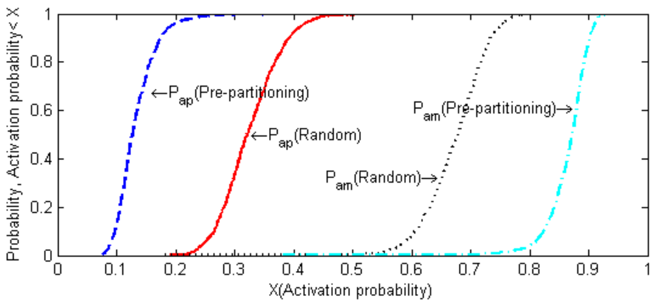

Figure 6 CDF of activation probability at discrete points over the service area

It can be seen that in the pre-partitioning scheme the probability of a user to be activated in preplay stage belongs to a certain range from 8% to 20%. The average activation probability in preplay from this point of view is 14%. Users will have an average activation probability of 35% if they have the access to the entire spectrum pool. Accordingly, the average activation probability of the pre-partitioning scheme in the main stage is 86%, compared with 65% in the basic scheme, an improvement of 21%. Hence, the reduction of activation in preplay is significant by applying pre-partitioning, yielding the significant benefits that are required.

4 Conclusions

In this paper a technique called spectrum pre-partitioning has been introduced, which is able to significantly reduce the exploration stage of a reinforcement learning cognitive radio. By randomly reserving a subset of available spectrum, the spectrum pool is fully partitioned before the transmission starts. Simulation results show that even if the cognitive users perform worse in preplay, the overall performance is still much better than the basic reinforcement learning scheme. By pre-partitioning the spectrum pool, about 86% of activations are accepted in the main stage. That is 21% higher than the scheme where a cognitive radio user is able to access all the available channels.

References

[1] Mitola, J. and J. Gerald q. Maguire, *Cognitive radio: making software radios more personal.* IEEE Personal Communication, Aug, 1999. **6**(4): p. 13-18.

[2] Haykin, S., *Cognitive Radio: Brain-Empowered Wireless Communications.* IEEE Journal on selected areas in communications, Feb. 2005 **23**: p. 201-220.

[3] Dasilva, L. and A. Mackenzie, *Cognitive Networks: Tutorial*, in *CrownCom* July 2007: Orlando, FL.

[4] Clancy, C., et al., *Applications of Machine Learning to Cognitive Radio Networks.* IEEE Wireless Communications, August 2007. **14**(4): p. 1536-1284.

[5] Leith, D.J. and P. Clifford. *Convergence of distributed learning algorithms for optimal wireless channel allocation*. in *45th IEEE Conf on Decision and Control*. 2006. San Diego.

[6] Bublin, M., et al. *Distributed spectrum sharing by reinforcement and game theory*. in *5th Karlsruhe workshop on software radio*. March. 2008. Karlsruhe, Germany.

[7] Tsagkaris, K., A. Katidiotis, and P. Demestichas, *Neural network-based learning schemes for cognitive radio systems.* Computer Communications, September 2008. **31**(14): p. 3394-3404

[8] Sutton, R.S. and A.G. Barto, *Reinforcement learning : an introduction*. 1998: The MIT Press.

[9] Jiang, T., D. Grace, and Y. Liu. *Performance of Cognitive Radio Reinforcement Spectrum Sharing Using Different Weighting Factors.* in *International Workshop on Cognitive Networks and Communications (COGCOM2008) in conjunction with CHINACOM'08.* August, 2008. Hangzhou, China.

[10] Jiang, T., D. Grace, and Y. Liu, *Cognitive Radio Spectrum Sharing Schemes with Reduced Spectrum Sensing Requirements*, in *The IET Seminar on Cognitive Radio and Software Defined Radios: Technologies and Techniques*. September, 2008: London.

[11] R.Saunders, S., *Antennas and propagation for wireless communication systems*. 1999: Wiley.